AN ADAPTIVE CELL SCHEDULING ALGORITHM FOR WIRELESS ASYNCHRONOUS TRANSFER MODE (ATM) SYSTEMS

FIELD OF THE INVENTION

This invention relates to the field of high speed wireless networks. More particularly, this invention relates to the field of integrating a scheduling algorithm to switch between contention and request and grant mode to achieve high performance in an asynchronous transfer mode (ATM) cell.

BACKGROUND OF THE INVENTION

The general structure of a high speed wireless network utilizes a central Hub and a number of End User Nodes (EUNs). The basic operating unit of the system is a cell, where the Hub is located at the center of the cell and the EUNs are distributed within the coverage range of the Hub. In the downstream direction (from the Hub to the EUNs), the communication channel is a broadcasting channel. In other words, any signal sent by the Hub is received by every EUN in the system.

In the upstream direction (from the EUNs to the Hub), each EUN transmission is a unicast heard only by the Hub. In a conventional high speed wireless network, the EUNs are not configured to receive transmissions from other EUNs. The upstream communication channel is shared by many EUNs using one of two types of access systems. A first access system is known as a time division multiple access (TDMA) based system is one in which all the EUNs transmit using the same frequency, but avoid colliding by transmitting at different times. A second access system is called a frequency division multiple access (FDMA) system is one in which each EUN has a separate frequency assignment to avoid collision when coordinating transmission. In both these access systems, the downstream signal is transmitted at a different carrier frequency than the upstream signal. However, because the upstream and downstream slots are of the same length, both channels have the same bandwidth.

As the EUNs are switched on, the Hub executes a one time ranging operation. Ranging determines the distance between each EUN and the Hub. An EUN which is far from the Hub must transmit faster than an EUN which is near to the Hub. After the

ranging is complete, the EUNs are accordingly aligned by the Hub in time slots from the frame of reference of the Hub.

During any given time slot, only one EUN is allowed to transmit at a time. To coordinate EUN transmission in a given time slot, the Hub may utilize one of two main coordinating modes.

A contention mode is utilized to coordinate transmission when there is a relatively small number of EUNs in the system trying to capture the upstream bandwidth. When there is a larger number of EUNs in the system, a request and grant, or piggybacking mode is used.

In the contention mode, the Hub broadcasts which EUNs are eligible to contend for a particular time slot. Each contending EUN will then answer the broadcast if that particular EUN wishes to transmit. This mode is particularly useful when there is a relatively small amount of EUNs in the system, or when idle EUNs require a route to request grants quickly.

The request and grant, or piggybacking mode functions oppositely of the contention mode. In the piggybacking mode, each individual EUN that wishes to transmit will unicast a request for a time slot from the Hub. The Hub will then broadcast a time slot grant signal to every EUN. When this occurs, each individual EUN will receive the signal and will recognize whether it, or another EUN is receiving a time slot grant. These grants will assign EUNs to time slots based on bandwidth availability. This mode is most effective when there are a large number of EUNs transmitting continuously.

Because the number of EUNs in a wireless system can change often and sometimes change in a short period of time, a scheduling algorithm is needed to adjust the ratio between request and grant mode and contention mode. To provide a smooth transition, an algorithm which can seamlessly switch between the two coordinating slotted multiple access modes is desired.

SUMMARY OF THE INVENTION

The present invention is a method of implementing an adaptive cell scheduling algorithm in a wireless network channel shared by a plurality of users. The present invention switches between two coordinating slotted multiple access modes to coordinate user transmission.

The two coordinating slotted multiple access modes, a contention mode and a request and grant mode, are both present in the network at all times and the percentage of each of the modes are dynamically changing values whose sum is always 100%.

The present invention allocates a queue in a weighted fair queue to generate contention slots. When a collision occurs between two users, two new requests for generating contention slots are placed in the weighted fair queue. When all contention is resolved, a starting request is placed in the weighted fair queue. This method allows the weighted fair queue to adjust the rate of generating contention slots automatically. When the network is lightly loaded, the weighted fair queue increases the rate of generating contention slots, and when the network is heavily loaded, the weighted fair queue decreases the rate of generating contention slots.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates an ATM cube in a stack configuration of the present invention.

FIG. 2 illustrates a Multicast ID Table of the present invention.

FIG. 3 illustrates the preferred Scheduling Block of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the preferred embodiment of the present method, the basis for transmission of broadband information in this type of high speed wireless network is asynchronous transfer mode (ATM). An ATM Cube may be considered in terms of a traditional stack configuration as depicted in Figure 1. More specifically, the Hub receives ATM cells from a back-haul network and delivers them in a special format to the EUNs.

In the reverse direction, an EUN receives ATM cells from its subscribers or converts legacy interfaces to ATM and delivers these ATM cells in a special format to the Hub. The Hub recovers the ATM cells from the upstream signals and delivers them to the back-haul network.

Any ATM system may be considered in terms of the traditional layered protocol stack (Fig. 1). In this framework, the ATM Layer(10) is responsible for delivery of data across multiple physical connections. The ATM Adaption Layers (AALs)(12) and applications operate above the ATM Layer(10). Below the Network Layer, the ATM specification refers to a Physical Layer(PHY)(18). An ATM's PHY(18) consists of the

Transmission Convergence (TC) Sub-Layer(14) and the Physical Medium (PM). In terms of the traditional stack model, the TC is the Data Link Layer (DLL)(16) and the PHY(18) combined. In the traditional stack model, the DLL(16) is typically broken down into two sub-layers, a Logical Link Control (LLC), which interfaces to the network, and a MAC which interfaces to the PHY(18).

Referring again to Figure 1, an ATM system is also divided into vertical planes, which provide similar services in all protocol layers. These planes are the control, user, and management planes. The User Plane(22) is responsible for carrying data, including control and management information, across those connections and for performing functions such as flow control. The Control Plane(20) is responsible for performing call control and connection control functions. The Management Plane(24) is responsible for resource allocation, error and fault reporting, operation and maintenance (OAM) information flows, and any parameters associated with a given layer. In addition, an overall Layer Management(26) spans all layers and is responsible for the coordination of the entire system.

The preferred embodiment of the present method assigns each physical EUN a unique 6-byte physical identification that is used during sign-on. Thereafter, it is assigned a logical EUN_ID from 1 to 254 (i.e. 0000_0001 to 1111_1110). The EUN_ID is changeable on the fly and a physical EUN may be assigned multiple EUN_IDs in order to allow it to have more than 256 Virtual Circuits (VCs). The link address is a 16 bit number, broken down into an 8 bit EUN_ID, with an 8 bit sub-address. These bits can be used to specify a null, unicast or multicast/broadcast addresses.

According to the preferred embodiment of the present invention as depicted in Figure 2, multicast addressing is used for sending control messages and to assign contention slots. The multicast addresses use an EUN_ID of 255 (1111_1111). This form of addressing uses the 8 bit sub-addresses to specify a group of EUNs. As shown in Figure 2, a group may consist of all EUNs down to a specific EUN.

The upstream air interface of the preferred embodiment uses TDMA. Individual upstream slots are assigned to a specific VC in a specific EUN. Contention timeslots are assigned to a group of EUNs, using the multicast addressing mode. Since there is no retry mechanism, only cell requests may be transmitted in contention slots.

Two requests are included in each upstream burst in the MAC messaging field. Each request includes an 8 bit link layer connection (VCI) address (i.e. a link address without the EUN_ID, which would be the same as that in the Link Address field), and an 8 bit length field, indicating the number of cells which are currently queued at the EUN. Because the actual length is transmitted, the effects of a lost or duplicated request are minimized.

The preferred embodiment of the invention also provides a contention mechanism to allow the EUN to notify the Hub of pending traffic. It is anticipated that the majority of the requests will be carried in normal MAC cells (piggybacking), but the contention mechanism provides a route for idle EUNs to request grants quickly. An upstream contention slot will be announced to all EUNs via a grant from the Hub using the multicast addressing mode, which indicates the group of EUNs that may contend.

The preferred embodiment of the present invention is depicted in Figure 3 and is configured to seamlessly switch between request and grant mode and contention mode. A scheduling algorithm incorporated by a scheduling block(30) containing a priority queue(32), a bypass queue(34), a weighted fair queue(36) and a contention queue(38) is used to implement this algorithm. The bypass queue(36) allows real time traffic such as Constant Bit Rate(CBR) to bypass the weighted fair queue(34) and not be included in the algorithm with the contention mode and the request and grant mode. Therefore, when the bypass queue(36) is not empty, the priority queue will allow the contents of the bypass queue(36) to be delivered before delivering the contents of the weighted fair queue(34).

Referring again to Figure 3, the weighted fair queue(36) actually implements the scheduling algorithm to serve non-real-time traffic classes so different streams such as EUN requests(40) and the output from the contention queue(38) can be prioritized based on their Quality of Service (QoS). The QoS is the priority given to any given EUN based upon the monetary value paid by that EUN to use the wireless network. For instance, an EUN that pays more in QoS will have a higher priority and therefore be higher in line than an EUN that pays less. The contention queue, however, has a priority lower than that of any EUN input(40) requesting bandwidth from the weighted fair queue(34).

Therefore, according to the preferred embodiment of the present invention, when the network is heavily loaded and a number of EUNs are requesting bandwidth(40) from the weighted fair queue(34), the contention queue(38) will occupy a small percentage of

the total requests to the weighted fair queue(34). For instance, if there are nine EUNs requesting bandwidth(40) from the weighted fair queue(34) in request and grant mode, the ratio of request and grant/contention will be 9/1 or 90% request and grant and 10% contention.

Referring again to Figure 3 in the preferred embodiment of the present invention, as the network becomes lightly loaded, the number of EUNs requesting bandwidth(40) from the weighted fair queue(34) decreases. However, the contention queue(38) retains the one input to the weighted fair queue(34). Therefore, as the number of EUNs requesting bandwidth(40) in the request and grant mode decreases, the ration of request and grant/contention decreases. For instance, if the number of EUNs requesting bandwidth(40) decreases to three, then the mode ration will be 3/1 or 75% request and grant and 25% contention. In other words, the percentage of contention in the system will seamlessly increase as the number of EUN requests(40) in the system decrease.

Alternative embodiments of the present invention include variations of the scheduler block(30) such as omissions, additions or reorganizations of the priority queue(32) or the bypass queue(36). However, the organization of the weighted fair queue(34), the contention queue(38) and the EUN requests(40) will not differ from the preferred embodiment.

Other aspects of the preferred embodiment of the present invention include the MAC continuously selecting cells to transmit in the downstream direction and assign upstream slots to specific VCs at specific EUNs. The downstream direction is straightforward, being very similar to existing wireline ATM systems. A Weighted Fair Queue (WFQ) is used for data traffic while the CBR traffic is transmitted immediately on a First In First Out (FIFO) basis.

Allocating the upstream bandwidth in the preferred embodiment is significantly more complicated. Each EUN sends requests to the Hub for specific VCs, which are carried with other upstream cells or vital to the contention mechanism. The Hub takes these requests and converts them into pseudo cells, which are input to an ATM scheduler. The Hub also predicts the arrival of CBR traffic, and generates pseudo cells for them automatically. To make each assignment, the Hub examines the virtual cell selected by the scheduler and grants an assignment to the corresponding VC at the appropriate EUN. Normally, requests are carried with other traffic in the request fields of upstream MAC

cells. However, a contention mechanism is provided to significantly reduce latency when an otherwise idle EUN receives cells to transmit upstream.

The sign-on procedure of the preferred embodiment is used by an EUN to establish initial ranging and power adjustments and establish a MAC level connection with the Hub and is a variation of the standard ranging process as described in [Survey of Ranging and Calibration Algorithms, KA]. Following sign-on, the DLL performs authentication of the newly connected EUN and VCs. A standard sign-off procedure is provided to allow the Hub or a local operator to gracefully terminate an EUN's connection. A mechanism for detecting connection loss due to extreme RF conditions and algorithms to automatically reestablish the connection is provided.

It will be readily apparent to one skilled in the art that other various modifications may be made to the preferred embodiment without departing from the spirit and scope of the invention as defined by the appended claims.